Effective components of social skills training programs for children and adolescents in nonclinical sample

de Mooij, B.; Fekkes, M.; Scholte, R. H. J.; Overbeek, G.

Published in:
Clinical Child and Family Psychology Review

DOI:
10.1007/s10567-019-00308-x

Publication date:
2020

Document Version
Publisher's PDF, also known as Version of record

Effective Components of Social Skills Training Programs for Children and Adolescents in Nonclinical Samples: A Multilevel Meta-analysis

de Mooij, B.; Fekkes, M.; Scholte, R.H.J.; Overbeek, G.

DOI
10.1007/s10567-019-00308-x

Publication date
2020

Document Version
Final published version

Published in
Clinical Child and Family Psychology Review

Clinical Child and Family Psychology Review (2020) 23:250–264
https://doi.org/10.1007/s10567-019-00308-x

Effective Components of Social Skills Training Programs for Children and Adolescents in Nonclinical Samples: A Multilevel Meta-analysis

Brechtje de Mooij[1] · Minne Fekkes[2] · Ron H. J. Scholte[3] · Geertjan Overbeek[1]

Published online: 9 January 2020

Abstract

Social skills training (SST) programs for nonclinical children and adolescents are known to have positive effects on social skills, but it remains unclear how distinct training components are related to program effects. This multilevel meta-analysis examines how psychoeducation (i.e., exercises aimed at the transfer of knowledge), psychophysical components (i.e., physical exercises aimed at improving self-confidence and trust in others), skill-building components (i.e., exercises aimed at improving interpersonal skills), and cognitive-emotional components (i.e., exercises aimed at changing emotions and cognitions) are independently related to SST program effects. We extracted data from 97 articles describing 839 effect sizes. Training content data were extracted from 60 corresponding SST programs. Our results showed that SST programs had a positive effect on the development of interpersonal skills and emotional skills in nonclinical samples: $d = .369$, 95% CI [.292, .447], $p < .001$. This effect was positively influenced by the inclusion of psychoeducation and skill-building components. The inclusion of psychophysical components and the number of cognitive-emotional components did not influence program effects. For psychoeducation and skill-building components, we observed a curvilinear relationship between intensity and effect size: programs including three to six psychoeducational exercises yielded larger effect sizes compared to programs with more or fewer psychoeducational exercises, and programs with 11 to 20 skill-building exercises outperformed programs with more or fewer skill-building exercises. These findings are an indication that psychoeducational components and skill-building components are related to larger SST program effects, granted that the dosage is right.

Keywords Meta-analysis · Social skills training programs · Effective elements · Intervention effectiveness · Training components

Shy and anxious children that are afraid of being laughed at by others are not fun playmates for their peers. These children may be excluded from activities, may become more socially withdrawn, and may even become the target of bullying behavior by others. On the other end of the spectrum are domineering, controlling children that become angry quickly, unable to regulate their impulses and emotions. These children are not fun playmates for their peers either and may also be at risk for marginalization in the peer context (Cook et al. 2010). The two types of children described above behave in very different ways, but both have difficulty in conducting themselves appropriately in social interaction. A social skills deficit can be a risk factor for different adverse outcomes, such as peer rejection and depression (Segrin 2000), antisocial problem behavior and delinquency (Ang 2003), and academic failure (Malecki and Elliot 2002).

Social skills can be viewed as a multidimensional construct that can be defined as learned behaviors that predict adaptive outcomes in social situations (Gresham and Elliot 1987). Being socially skilled reflects the ability to perform a variety of social behaviors adequately, such as problem-solving, assertion, cooperation, attribution, communication, emotional sensitivity, and emotion regulation (Kavale and

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10567-019-00308-x) contains supplementary material, which is available to authorized users.

✉ Brechtje de Mooij
L.S.deMooij@uva.nl

[1] Research Institute of Child Development and Education, University of Amsterdam, Nieuwe Achtergracht 127, 1018 WS Amsterdam, The Netherlands

[2] Child Health, TNO, Leiden 2316 ZL, The Netherlands

[3] Behavioural Science Institute, Radboud University Nijmegen, Nijmegen 6500 HE, The Netherlands

Forness 1996; Spence 2003). Having adaptive social skills is related to being perceived by others as socially competent (McCelland and Scalzo 2006) and to higher peer acceptance (Caprara et al. 2000). Social skills contribute to an individual's ability to initiate and maintain positive social interactions. For example, children that can adequately solve a conflict with their peers are flexible in choosing how to react in social situations, which can decrease impulsiveness and frustration with others (Denham and Almeida 1987).

Social skills training (SST) programs are those programs that aim to teach and improve children's appropriate, adaptive social behaviors. Over the past decades, many SST programs have been developed for both clinical and nonclinical target populations (e.g., Frey et al. 2015). These SST programs are generally multifaceted, stacking different training components, such as psychoeducation, skills training, and cognitive-behavioral exercises. Consequently, SST programs typically target multiple outcomes.

SST programs can be based on multiple theory-based approaches on how children learn social skills. Social learning theory holds that social behavior is learned by observing others (Bandura 1978). Consequently, many SST program components focus on modeling, coaching, and shaping behavior (Ladd and Mize 1983). The reinforcement theory posits that behavior, both positive and negative, is more likely to occur when a reward follows it. Based on this approach, several SST program components focus on improving children's social competence by directly and explicitly rewarding prosocial behavior (Skinner 1953). Furthermore, the cognitive-behavioral approach highlights the importance of interpersonal cognitive problem-solving skills. This approach results in SST program components that focus on how to deal with others in alternative ways through means-end thinking (Denham and Almeida 1987).

Despite having a common aim, SST programs vary widely in content, design, and target population (Diekstra 2008). For example, a program can be embedded within the school curriculum or be self-contained, and can target specific behavior (e.g., bullying or social anxiety) or numerous different behaviors. Also, it can target children with clinical levels of behavioral problems, children at risk for behavioral problems, or children from the general population. SST programs also vary widely in their focus on stimulating either one specific skill or a combination of skills, including interpersonal problem-solving, more adequate processing of social information, adjusting cognitive distortions, increasing social knowledge, increasing self-regulation, and the acquisition of an appropriate set of social behaviors (Spence 2003).

The effectiveness of SST programs has been studied extensively, and several meta-analyses assessing the effects of SST programs in nonclinical and at-risk populations have been published. However, mixed findings have emerged regarding the effectiveness of SST programs on different outcomes. Previous meta-analyses have assessed universal prevention and intervention programs and have generally shown positive effects on direct outcomes of SST for children and adolescents. SST programs have shown effect sizes on problem-solving skills that range from $d = .26$ to .78, effect sizes for SST on self-esteem range from $d = .16$ to .69, effect sizes for SST on social competence range from $d = .22$ to .70, and effect sizes for SST on social behavior range from $d = .24$ to .92. Program effects on secondary outcomes are slightly smaller: effects on disruptive behavior and aggression range from $d = .12$ to .63 and $d = .12$ to .40, respectively (Denham and Almeida 1987; Diekstra 2008; Durlak et al. 2011; Lösel and Beelmann 2003; Reddy et al. 2009; Schneider 1992; Sklad et al. 2012; Taylor et al. 2017).

Based on previous research, we know that SST programs work, but we do not know to what extent program effects depend on individual training components (Chorpita and Daleiden 2009). Most of the SST programs included in meta-analytic studies combine different program components, which is why several scholars have compared these types of multifaceted training programs to cocktails (Leijten et al. 2015). Until now, research has predominantly focused on the cocktail as a whole; the effectiveness of individual training components has mostly gone untested. This hampers our insight into how distinct training components are related to program effects, limiting practitioners' ability to tailor SST programs to their client needs. Such knowledge about effective training components can be used to adjust existing SST programs to exclude components that do not stimulate positive child development, or that may even produce iatrogenic intervention effects (Dishion et al. 1999). Therefore, we conducted a meta-analysis to determine if and how distinct training components are related to SST program effects for children and adolescents in a nonclinical sample. This allowed us to investigate the specificity hypothesis described by Chorpita and Daleiden (2009), which proposes that specific components are related to intervention effects.

The Present Meta-analysis

Using meta-analytic strategies, we are the first to examine the associations of distinct training components with effect sizes of SST programs. There were two reasons for our focus on a nonclinical target audience. First, SST programs are widely implemented in schools, mostly reaching children with light or just emerging problem behavior. The second reason was a practical one: as there is a wide variety of SST programs for different target populations, it was not feasible to include both clinical and nonclinical samples. This study was a first attempt to relate SST components to SST

program effects, and therefore, we demarcated the scope of our meta-analysis.

We disassembled and coded 60 SST programs in terms of their distinct training components. These training components were then related to the effect sizes of 97 randomized controlled trials and controlled quasi-experimental studies using a multilevel approach. This enabled us to investigate whether specific training components of SST programs are related to larger or smaller effect sizes on primary and secondary outcomes. Evidence points to social skills deficits underlying the development of problem behavior (e.g., Spence 2003). Therefore, the main analyses of this study are focused on the effects of SST programs on the primary outcomes of interpersonal and emotional skills. For nonclinical samples, these are the outcomes that are most directly targeted in SST programs. Effectively targeting interpersonal and emotional skills in SST programs should prevent the development of more serious problem behaviors. Therefore, we assessed SST program effects on problem behavior outcomes in the secondary analyses.

This meta-analysis is the first scientific endeavor to connect distinct training components to the effects of SST programs for nonclinical children and adolescents and is thus exploratory in nature. Following the specificity hypothesis, we expected that SST training components would be differently associated with SST program effects. However, we did not formulate a specific hypothesis about which training components would be more or less effective.

Method

For this study, we followed the Preferred Reporting Items for Systematic reviews and Meta-Analyses guidelines (Moher et al. 2009) and the guidelines issued by the American Psychological Association (2010). Figure 1 provides an overview of the flow of information through the different phases of our study.

Information Sources

The electronic search was conducted in multiple databases: PsychINFO, ERIC, Medline, Scopus, and Google Scholar. The search filter consisted of five elements containing key terms for (1) the primary focus of the study, (2) the secondary focus of the study, (3) the type of training, (4) the target audience, and (5) the type of study. The element for the primary focus of the study contained the search words: social skills, social skills training, social*, socio*, interpersonal, intrapersonal, skills, interpersonal competenc*, intrapersonal competenc*, social competenc*, soci* emotional learning, and peer relations. The element for the secondary focus of the study contained the search words: assertiveness, test anxiety, performance anxiety, emotional control, anger control, prosocial behavior, assertive*, coping strateg*, resilience, and prosocial. The full search string is available in Online Appendix A.

Selection of Studies

Data collection was twofold as data were obtained from both published studies and program manuals. The selection of the articles, as well as the retrieval of the program manuals, was executed following a prewritten protocol. This protocol is available in Online Appendix B.

We defined "SST program" as a program aimed at teaching or developing children's adaptive social behavior to improve their success in social interactions. Studies eligible for review (i) assessed the effectiveness of an SST program for school-age children and adolescents up to 18 years of age, (ii) assessed an SST program that targets a sub-clinical population, (iii) had a pre-test post-test design with a control group, (iv) reported at least one social skill outcome, (v) provided sufficient statistics to calculate a Cohen's *d* effect size, (vi) were written in English or Dutch, (vii) were peer-reviewed, and (viii) were published from January 1, 1990, onwards. We did not include studies published before 1990 to ensure that SST program manuals would still be available, and the included studies reflected relatively recent evidence for SST program effectiveness.

We excluded studies if the training assessed was a parenting program aimed at changing children's behavior by teaching parents how to discipline or interact with their children. Programs and studies that focused on children's or adolescents' physical health (e.g., prevention of drug use, AIDS, pregnancy, and so on) were also excluded. Moreover, programs focusing on preschool-aged children were excluded, as well as studies that assessed SST programs in children with attention deficit hyperactivity disorder (ADHD), autism, or clinical levels of internalizing or externalizing behavior. We did not formulate criteria concerning the SST program setting or type. Thus, school-based and community-based programs as well as universal and indicated SST programs were eligible for this study.

Selection of Training Programs

In total, 6206 eligible records were obtained. The titles and abstracts of all of these records were screened for inclusion criteria. Based on our exclusion criteria (see above, "Selection of Studies" paragraph), 5836 of the 6206 records were excluded in the identification phase. The remaining 370 eligible articles met our inclusion criteria based on the screening of the title and abstract. These articles corresponded to 188 SST programs. Efforts to obtain program manuals started in September 2016. Both the study authors

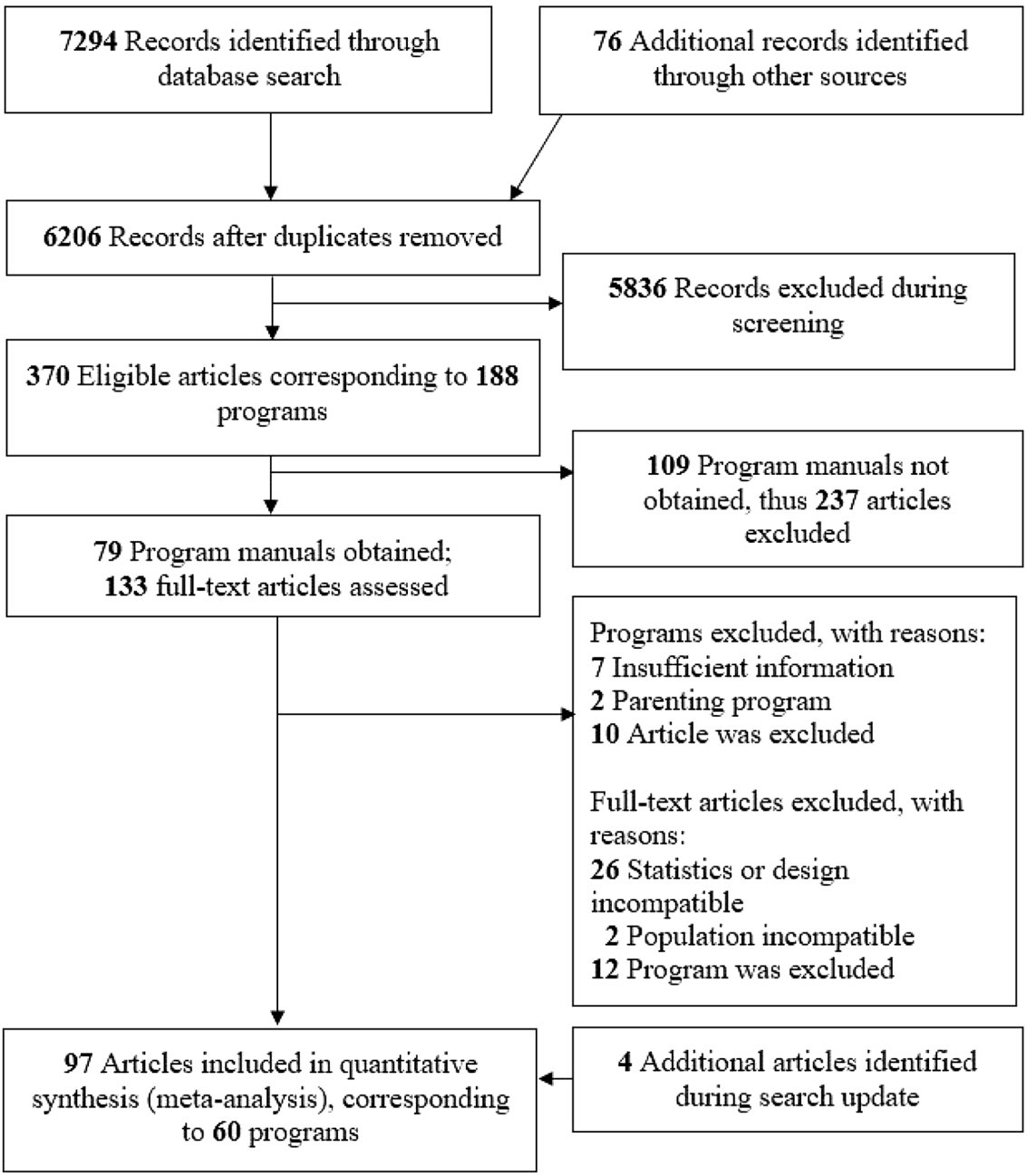

Fig. 1 PRISMA flowchart of study and program selection process

and program developers of the 188 eligible SST programs were contacted to request a copy of the program manual. We could not obtain a copy of 109 programs, which resulted in the exclusion of 237 articles. Efforts to obtain program manuals ended May 31, 2017. By this time, the manuals of 79 SST programs had been collected.

The 79 SST programs corresponded to 133 articles, of which the full text was assessed in the eligibility phase of our study. After reading the complete articles, another 28 articles were excluded due to a lack of statistics, or the design or population not meeting our inclusion criteria. The 10 SST program manuals that corresponded to these articles were, therefore, also excluded. After a full inspection of the manuals, nine SST programs were excluded. Seven of these programs did not provide enough information in the program manual to allow for reliable coding of the exercises. The two other programs were excluded because they focused on parents instead of children. The 12 articles that corresponded to these excluded SST programs were consequently also excluded from our study. Our search was updated in October

2018, adding four studies that described previously coded SST programs. Ultimately, 60 unique SST programs were included. These SST programs were evaluated in 97 articles that reported on $N = 71{,}226$ participants between 3 and 17 years old. We extracted a total of 839 effect sizes from the included studies (see Fig. 1 for the full study flowchart; Liberati et al. 2009). Online Appendix C provides references for the included studies.

Of the included training programs, 19 were social-emotional learning (SEL) programs. Other programs targeted more specific behavior: six programs targeted bullying behavior, ten programs targeted (social) anxiety, 11 programs targeted disruptive behavior, five programs targeted resilience and self-esteem, and nine programs targeted prosocial interactions. Online Appendix D, Table D.1, provides an overview of the included programs and studies.

Data Extraction and Coding

After all eligible studies and corresponding program manuals were collected, data were coded using two separate coding systems, one for coding the studies and one for coding the program manuals. We did not require SST programs to be available in English or Dutch, because we did not want to exclude possibly effective programs that were not (yet) translated into English from our study. If program developers or authors consented to share the program manual with us, but the program manual was not available in English or Dutch, all exercises included in the program were discussed during a Skype meeting. In these Skype meetings, we used the coding system as a guide to the semi-structured interview and asked program developers or authors to explain every exercise of the program, which allowed us to code the program from the information provided verbally. In this way, we safeguarded the reliability of the coding process and made sure that all included programs were scored on the same criteria. Six programs were coded this way.

To code the program manuals, we developed a taxonomy based on a previous taxonomy by Veerman et al. (2015), supplemented with components from other international taxonomies (Chorpita and Daleiden 2009; Michie et al. 2013). All exercises of the program were categorized into one of four main training component categories or the booster component category. All categories of training components included subcategories, and these included up to eight training subcategories each.

The first of the main training component categories is *Psychoeducation*, which included all exercises that are aimed at transferring knowledge about behavior or group processes and social roles. The second category is *Psychophysical components,* which included physical relaxation exercises, physical exercises to improve posture, physical exercises aimed at boosting self-confidence and physical exercises aimed at promoting trust in others. The third category *Skill-building components* included exercises in verbal and non-verbal communication, teamwork exercises, exercises to promote and practice prosocial behavior, and exercises to improve problem-solving. The fourth category *Cognitive-emotional components* included exercises to enlarge self-awareness, exercises to practice recognizing one's own and others' emotions, exercises to improve impulse regulation, cognitive restructuring exercises, and mindfulness exercises. Finally, the category *Booster components* included program units aimed at class management, rewarding of behavior, behavioral contracting, and generalization to situations transcending the training, coaching, and (self-)monitoring behavior. The individual exercises were coded to belong to one of the five component categories exclusively. To assess the dosage with which the different components were implemented, we calculated the number of exercises per component category. Online Appendix D, Table D.2, provides an overview of the individual training components used in the featured SST programs.

Additionally, characteristics of the SST program were coded. We coded whether a program was universal (targets general population) or indicated (targets children with emerging behavioral problems), the duration of the program (in weeks), the type of trainer (school personnel, mental health professional or non-school personnel), the mode of delivery (computer program or face-to-face), and the age of participants (primary school age, secondary school age or both children and adolescents). If needed, this information was supplemented with information provided in the corresponding studies.

We also coded several study characteristics: bibliographic information (e.g., authors, journal of publication, year of publication, and location of study), sample size, the time between pre-test and first reported post-test (less than 6 months, 6 to 12 months, 13 months to 2 years, or more than 2 years) and the informant per effect size (self-report, behavior rating by others or behavioral observation). The quality of the study was assessed using the Quality Assessment Tool for Quantitative Studies (QATQS). The QATQS is a widely used, reliable tool to assess the quality of a study based on selection bias, study design, possible confounders, blinding, data collection methods, and withdrawals and dropouts. Each section is rated on a three-point scale (weak, moderate, or strong). The global rating for a study is determined by assessing the rating of the five sections. A study was rated as strong if there were no weak ratings on any of the domains. A study was considered moderate if a weak rating was assigned to one of the domains, and a study was considered weak when two or more domains were rated weak (Effective Public Health Practice Project 1998; Thomas et al. 2004).

The first author trained research assistants to code studies and program manuals. The coding schemes were discussed

extensively, and research assistants practiced coding a program manual that was not included in the final meta-analysis. The training was completed when there was sufficient agreement between coders. Ten studies were double coded by the first author and research assistants (10% of the total number of studies) to determine inter-rater agreement. Intraclass correlations (ICC) ranged between .664 and 1.00. For the manuals, inter-rater agreement was based on double coding of six program manuals (10% of the total number of manuals), and there was an agreement of 72.3% between raters (ICC = .684). The disagreement between raters could mostly be attributed to difficulty distinguishing psychoeducational exercises aimed at the transfer of knowledge only and psychoeducational introductions to other components—which often include a short introduction of behavior followed by an opportunity to practice. Disregarding all exercises coded as psychoeducational by either rater, the inter-rater agreement was 86% (ICC = .920). All disagreements between raters were reevaluated and solved through mutual discussion. Exercise codes were corrected for subsequent analyses.

Calculation of Effect Sizes

Only the effects of SST programs from pre-test to the first reported post-test were included in the current meta-analysis because not all of the included studies presented follow-up data. Raw data (means and standard deviations) were converted into Cohen's *d* values by calculating the mean difference between pre-test and first reported post-test of the experimental and control group, and dividing by the pooled standard deviation of the experimental and control group at pre-test (Morris 2008, Eq. 8). We used the pooled standard deviation to estimate the sampling variance more accurately, and the estimated effect size was adjusted according to sample size (Morris 2008, Eq. 10).

When raw data were not sufficiently reported, alternative statistics were used to calculate the effect size, such as *F*-test, *t* test, Mann–Whitney *U*, odds ratio, or regression coefficients. If a result was not significant, and consequently, statistics were not presented in an article, an effect size of zero was coded (Durlak and Lipsey 1991). Effect sizes were coded in such a manner that a positive effect size reflected a positive behavioral change (e.g., increase of social skills), and a negative effect size reflects a negative behavioral change (e.g., a decrease of self-control).

Relevant effect sizes were categorized by the outcome domain. The first two categories, interpersonal and emotional skills, reflected outcomes at the core of the construct social skills. *Interpersonal skills* included outcomes reflecting problem-solving, coping, social skills, assertiveness, and social competence. *Emotional skills* included outcomes reflecting self-efficacy, self-awareness, self-esteem, empathy, emotion regulation, and emotion knowledge. We also coded more distal outcomes, as these outcomes can be viewed as proxies for social skills deficits. The category *Peer relationship problems* included outcomes reflecting victimization and bullying perpetration. *Internalizing problem behavior* included outcomes reflecting depression, loneliness, and (social) anxiety, and *Externalizing problem behavior* included outcomes reflecting aggression, conduct problems, attention problems, and hyperactivity.

Statistical Analyses

Preparations for analyses included centering continuous variables on the mean and recoding categorical variables into dummies. Effect sizes were checked for outliers by screening for *z*-scores higher than 3.29 or lower than −3.29. Effect sizes exceeding these *z*-scores were manually brought back to the highest value within a *z*-score of ±3.29 (Tabachnick and Fidell 2007).

The random-effects multilevel analyses were conducted using the metafor-package (Viechtbauer 2010) in R (version 3.3.4). The advantage of using a random-effects model over a fixed-effects model is that it takes into account that observed effect sizes might vary from true effect sizes due to external factors. The advantage of the multilevel approach over a traditional univariate approach is that all relevant outcomes can be included in the analysis, without the need to aggregate effect sizes per study. The multilevel approach thus preserves information while achieving maximum power. The multilevel approach accounts for the nesting of effect sizes within studies (van den Noortgate and Onghena 2003). This is important because studies on the effects of SST programs vary broadly in design and hence are not homogeneous. We accounted for the uncertainty in estimating residual heterogeneity by applying the Knapp and Hartung (2003) adjustment, which reduces Type I error.

First, we estimated the overall effect of SST programs on the separate outcome domains using random three-level univariate models (for a step-by-step tutorial, see Assink and Wibbelink 2016). For the subsequent analyses, we used a subset containing the effect sizes for interpersonal and emotional skills only, since these were considered the proximal outcomes. We examined the overall effect of SST programs for these two outcome domains together, and we assessed whether programs' inclusion of distinct training components was associated with stronger effects for SST programs on these proximal outcomes. A significant *Q*-test of heterogeneity indicates that a component significantly influences the overall effect size (Borenstein et al. 2009). In this phase of the analyses, we also examined if the total amount of exercises focused on the specific training components was associated with stronger effects for SST programs. Both continuous and categorical variables were included in the analyses as moderators. The categorical variable for the total amount

of exercises was formed based on the distribution of the continuous variable, in such a manner that each category of the variable included approximately 20% of the effect sizes. The mean estimated effect sizes per category of the dosage variable were compared to each other, to assess whether significant differences were present between the subgroups formed for the total amount of exercises per component category. Third, the moderating effect of booster components (i.e., the use of rewards, goal setting, generalization, coaching, and self-monitoring) was assessed. Then, we assessed if the coded program characteristics and research design characteristics were significant moderators of SST program effects.

We used the PET-PEESE approach to assess publication bias (Stanley and Doucouliagos 2014). This approach uses a meta-regression based model. The first step of this approach is the precision-effect test (PET), which is based on Egger's test and uses the standard error as a moderator of effect size. This test examines whether there is a true effect beyond publication bias; a significant moderator effect of the standard errors implies the presence of publication bias. When the intercept in the PET model is significant, a precision-effect estimate with standard error (PEESE) test is assessed. This test uses the variance of effect sizes as a moderator of effect size. In this second step, a significant variance of the effect size with the standard error implies publication bias (Stanley and Doucouliagos 2014). Additionally, we examined the symmetry of a funnel plot. An asymmetrical funnel plot indicates bias (Borenstein et al. 2009).

In secondary analyses, the influence of training components was separately assessed for peer relationship problems, internalizing behavior problems, and externalizing problem behavior. For these secondary analyses, we also estimated mean effect sizes by type of SST programs for each different secondary outcome domain.

Results

Effects of Individual Training Components on Interpersonal and Emotional Skills

SST programs yielded significant, small overall effects on all outcome domains. The largest effects were found for interpersonal and emotional skills (Table 1). As SST programs target interpersonal and emotional skills directly, and problem behavior outcomes are assessed as a proxy for improvements in interpersonal and emotional skills, our primary analyses focused on SST component effects on interpersonal and emotional skills. These analyses were based on data from 49 SST programs and $k = 77$ studies that reported on a total of 369 effect sizes. A random three-level meta-analysis yielded a significant, small overall SST program effect on interpersonal and emotional skills: $d = .369$, 95% CI [.292, .447], $p < .001$ (Durlak 2009). This effect translates to an average percentile gain of 13% on interpersonal and emotional skills following an SST (Coe 2002; McCartney and Rosenthal 2000). The analyses also demonstrated that there was significant heterogeneity in the effects of SST programs on interpersonal and emotional skills. Specifically, log-likelihood tests showed significant variance between effect sizes within studies ($\sigma^2_{\text{level 2}} = .053$; $\chi^2 = 873.000$, $p < .001$) and between studies ($\sigma^2_{\text{level 3}} = .098$; $\chi^2 = 111.951$, $p < .001$). About 4% of the total variance could be attributed to within-study sampling variance (level 1), about 34% to differences between effect sizes within studies (level 2), and about 62% to differences between studies (level 3).

We assessed if individual training components were related to the SST program effect sizes (Table 2). The inclusion of psychophysical components was not associated with stronger effects of SST programs on interpersonal and emotional skills. The effects of SST programs were not moderated by the total number of included psychophysical exercises either. Whether the inclusion of cognitive-emotional components was related to SST program effects could not be assessed, as only one included SST program did not include any cognitive-emotional exercises, and therefore, an

Table 1 The effect of SST programs on the different outcome domains, and heterogeneity between and within studies

Outcome measures	*#p*	*#k*	*#ES*	Mean *d*	95% CI	*p*	$\sigma^2_{\text{level 2}}$	$\sigma^2_{\text{level 3}}$
Interpersonal skills	42	68	239	.386	.288–.484	< .001	.068***	.136***
Emotional skills	26	38	130	.328	.225–.431	< .001	.007***	.098***
Peer relationship problems	19	27	56	.255	.095–.415	.002	.063***	.133***
Internalizing problem behavior	35	52	182	.233	.159–.306	< .001	.028***	.047***
Externalizing problem behavior	39	60	232	.172	.078–.266	< .001	.022***	.127***

#p number of SST programs, *#k* number of studies, *#ES* number of effect sizes, *Mean d* mean effect size (*d*), *95% CI* confidence interval, $\sigma^2_{\text{level 2}}$ variance within studies, $\sigma^2_{\text{level 3}}$ variance between studies

*$p < .05$, **$p < .01$, ***$p < .001$

Table 2 Results of moderator analyses of training components on interpersonal and emotional skills using a three-way univariate model

Moderator	Category	*#p*	*#k*	*#ES*	Estimate	95% CI	*Q*	*p*
Psychoeducation	Not included in the program	10	14	96	.181*	.014–.348	F (1, 367) = 6.026	.015
	Included in the program	39	63	273	.415***	.331–.499		
Total number of exercises	None	10	14	96	.181*	.017–.345	F (4, 361) = 2.253	.063
	1–2 exercises	10	14	61	.421***	.238–.604		
	3–6 exercises	10	15	39	.537***	.344–.729		
	7–14 exercises	9	12	79	.408***	.240–.575		
	15 > exercises	10	20	91	.312***	.168–.455		
Psychophysical components	Not included in the program	20	57	151	.388***	.275–.501	F (1, 367) = .187	.655
	Included in the program	29	20	218	.353***	.246–.461		
Total number of exercises	None	21	37	151	.387***	.275–.499	F (3, 357) = .574	.633
	1 exercise	12	19	94	.399***	.240–.559		
	2–4 exercises	11	15	48	.328***	.141–.516		
	5 > exercises	3	3	68	.215	-.050–.480		
Skill-building components	Not included in the program	3	3	7	.314	-.110–.739	F (1, 367) = .068	.794
	Included in the program	46	74	362	.372***	.292–.451		
Total number of exercises	None	3	3	7	.314	-.104–.732	F (4, 348) = 1.448	.218
	1–10 exercises	13	17	107	.282***	.125–.439		
	11–20 exercises	15	23	102	.499***	.359–.639		
	21–30 exercises	9	20	89	.284***	.127–.440		
	30 > exercises	6	8	48	.393***	.172–.613		
Cognitive-emotional components	Not included in the program	1	1	1	–	–	–	–
	Included in the program	48	76	368	–	–		
Total number of exercises	None	1	1	1	.259	-.568–1.087	F (5, 347) = .312	.905
	1–5 exercises	9	12	103	.376***	.191–.560		
	6–10 exercises	12	13	66	.447***	.261–.633		
	11–15 exercises	7	10	28	.407**	.157–.657		
	16–20 exercises	9	17	79	.362***	.202–.522		
	21 > exercises	8	18	76	.299***	.135–.463		

#p number of SST programs, *#k* number of studies, *#ES* number of effect sizes, *Estimate* estimate of effect size, *95% CI* confidence interval, *Q* *Q*-test of heterogeneity

*$p < .05$, **$p < .01$, ***$p < .001$

adequate comparison could not be made. Assessment of the total amount of cognitive-emotional components included in an SST program showed that this was not a significant moderator of SST program effects.

The inclusion of psychoeducation was significantly related to the effectiveness of SST programs on interpersonal and emotional skills. SST programs that included psychoeducation yielded a significantly larger estimated effect compared to programs that did not include this component, $\beta = .234$, $p < .05$. We did not find a linear association between the total number of psychoeducational exercises and SST program effects. Therefore, we performed a category comparison, which showed that programs with three to six psychoeducational exercises yielded a significantly larger estimated effect size, F (1, 364) = 3.893, $p < .05$, $d = .538$, $p < .001$, compared to programs not in this category, $d = .325$, $p < .001$ $\beta = -.213$, $p < .05$. In other words, there was a curvilinear relation, programs with three to six psychoeducational exercises outperformed programs with fewer as well as more psychoeducational exercises.

At first sight, the inclusion of the skill-building component did not appear to influence the effects of SST programs on interpersonal and emotional skills. However, a category comparison on the number of skill-building exercises showed that SST programs containing 11 to 20 exercises aimed at this component yielded a significantly larger estimated effect size, F (1, 351) = 5.152 $p < .05$, $d = .497$, $p < .001$, compared to programs not in this category, $d = .305$, $p < .001$, $\beta = -.193$, $p < .05$. Programs containing 11 to 20 skill-building exercises outperformed programs with fewer as well as more skill-building exercises.

Finally, we assessed if the inclusion of booster components accounted for variance in effect sizes. None of the booster components (i.e., class management, generalization,

rewarding of behavior, coaching, goal setting, and [self-] monitoring) were significantly related to SST program effects on interpersonal and emotional skills (Table 3). Thus, these booster components do not independently influence SST program effects. Also, none of the variables for the dosage of booster components were significant.

Program Characteristics

Table 4 presents all the results for the moderator analyses with SST program characteristics. Indicated SST program effects on interpersonal and emotional skills were not significantly different from universal SST program effects. Furthermore, there was no linear relation between program duration and program effects. However, a category comparison showed that SST programs of 27 weeks or more were found to have a significantly smaller estimated effect on interpersonal and emotional skills F (1, 361) = 4.567, $p < .05$, $d = .208$, $p < .001$ compared to programs of 10 to 11 weeks, $\beta = -.244$, $p < .05$, and programs of 12 to 16 weeks, $\beta = -.302$, $p < .05$.

Moreover, the type of trainer providing the program did not influence SST program effects on interpersonal and emotional skills. Noteworthy, however, is that the mean effect size of SST programs provided by non-school personnel (e.g., research staff or students) was not significant, whereas the mean effect size of SST programs provided by mental health professionals or school personnel was significant. It made no difference if a program required a pre-intervention training of certification for trainers. Furthermore, computer programs and face-to-face programs both yielded equally positive results, and the age of the participants did not influence SST effects either. These results are presented in Table 4.

Robustness of Main Findings

Research Design Characteristics

Analyses showed that the quality of the study significantly influenced the estimated mean effect of SST programs on interpersonal and emotional skills, F (2, 366) = 9.243, $p < .001$. Studies of moderate and strong quality yielded smaller effects, $d = .455$, $p < .001$ and $d = .167$, $p < .01$, respectively, compared to studies of weak quality, $d = .534$, $p < .001$. Also, the sample size of a study significantly influenced the estimated mean effect size, F (2, 367) = 9.464, $p < .01$. As the sample size of a study increased, the estimated mean effect of SST programs decreased, $d = .431$, $p < .001$, $\beta = -.0001$, $p < .01$. The time between pre-test and first reported post-test and the type of informant were not significant moderators of the effect of SST programs on interpersonal and emotional skills, F (2, 361) = 1.397,

Table 3 Results of moderator analyses of booster components on interpersonal and emotional skills using a three-way univariate model

Moderator	Category	*#p*	*#k*	*#ES*	Estimate	95% CI	*Q*	*p*
Class management	Not included in the program	4	8	34	.322*	.069–.575	F (1, 367) = .155	.694
	Included in the program	45	69	335	.375***	.293–.457		
Total number of exercises		46	69	353	.370***	.291–.450	F (1, 351) = 1.436	.232
Rewarding	Not included in the program	30	52	229	.404***	.310–.497	F (1, 367) = 1.687	.195
	Included in the program	19	25	140	.295***	.159–.431		
Total number of exercises		47	74	361	.364***	.286–.443	F (1, 359) = .281	.596
Goal setting	Not included in the program	36	62	263	.379***	.291–.468	F (1, 367) = .213	.644
	Included in the program	13	15	106	.335***	.170–.501		
Total number of exercises		48	76	364	.368***	.288–.447	F (1, 362) = .878	.349
Generalization	Not included in the program	15	26	115	.374***	.234–.513	F (1, 367) = .004	.948
	Included in the program	34	51	254	.368***	.274–.462		
Total number of exercises		47	74	361	.371***	.293–.451	F (1, 359) = 1.962	.162
Coaching	Not included in the program	12	16	78	.377***	.214–.539	F (1, 369) = .010	.922
	Included in the program	37	61	291	.368***	.279–.457		
Total number of exercises		46	71	353	.370***	.290–.450	F (1, 351) = 1.169	.280
(Self-)monitoring	Not included in the program	36	61	326	.383***	.296–.471	F (1, 367) = .464	.496
	Included in the program	13	16	43	.316***	.141–.490		
Total number of exercises		47	72	357	.382***	.296–.468	F (1, 355) = .001	.971

#p number of SST programs, *#k* number of studies, *#ES* number of effect sizes, *Estimate* estimate of effect size, *95% CI* confidence interval, *Q* *Q*-test of heterogeneity

*$p < .05$, **$p < .01$, ***$p < .001$

Table 4 Results of moderator analyses of program characteristics for the effect of SST programs on interpersonal and emotional skills using a three-way univariate model

Moderator	Category	*#p*	*#k*	*#ES*	Estimate	95% CI	*Q*	*p*
Setting	Universal program	37	64	325	.372***	.288–.457	F (1, 367) = .022	.883
	Indicated program	12	13	44	.356**	.155–.556		
Duration of the program	1–9 weeks	9	11	94	.275**	.091–.460	F (4, 358) = 2.033	.089
	10–11 weeks	14	21	94	.452***	.313–.592		
	12–16 weeks	8	13	36	.510***	.312–.708		
	17–26 weeks	8	14	77	.376***	.199–.552		
	27 > weeks	7	15	62	.208*	.046–.370		
Type of trainer	School personnel	33	56	299	.343***	.255–.431	F (2, 359) = .351	.704
	Mental health professional	13	15	56	.426***	.246–.605		
	Non-school personnel	3	3	7	.403	−.008–.813		
Schooling required for trainer	Yes	29	51	254	.377***	.281–.473	F (2, 366) = .221	.802
	No	12	16	76	.390***	.217–.564		
	Not specified	8	10	39	.305**	.093–.516		
Mode of delivery	Computer program	3	3	6	.525*	.079–.972	F (1, 367) = .485	.487
	Face-to-face	46	74	363	.365***	.286–.444		
Age of participants	Primary school age	31	50	268	.391***	.295–.486	F (2, 366) = .573	.565
	Secondary school age	7	7	21	.428**	.136–.720		
	Children and adolescents	11	20	80	.298**	.142–.455		

#p number of SST programs, *#k* number of studies, *#ES* number of effect sizes, *Estimate* estimate of effect size, *95% CI* = confidence interval, *Q* *Q*-test of heterogeneity

*$p < .05$, **$p < .01$, ***$p < .001$

$p = .243$ and F (2, 354) = .886, $p = .413$, respectively. See Online Appendix E, Table E.1 for the full results.

Publication Bias

Publication bias was assessed using the PET-PEESE method (Stanley and Doucouliagos 2014). The PET result showed that the standard error of effect sizes was a significant moderator of the effect size. As the intercept in the PET model was significant, the PEESE model was also assessed. The PEESE model, which includes variance as a moderator, was significant, and this implies publication bias. A funnel plot with the effect sizes on the *X*-axis and the standard error of effect sizes on the *Y*-axis (Fig. 2) shows that there is missing data on the right side of the funnel. This means that there are relatively few studies with larger sample sizes that report large positive effects (Duval and Tweedie 2000).

Outliers

Fifteen effect sizes were considered outliers (z-score exceeded ± 3.29). Outliers were manually adjusted to values within a z-score of ± 3.29 ($d = -1.30$ and $d = 2.04$, respectively). To assess whether this adjustment changed the results, we repeated the analysis of the overall effect with the unadjusted effect sizes. The overall estimated effect with unadjusted effect sizes was $d = .383$, 95% CI [.300, .466], $p < .001$, which is comparable to the effect originally estimated with the adjusted effect sizes, $d = .369$.

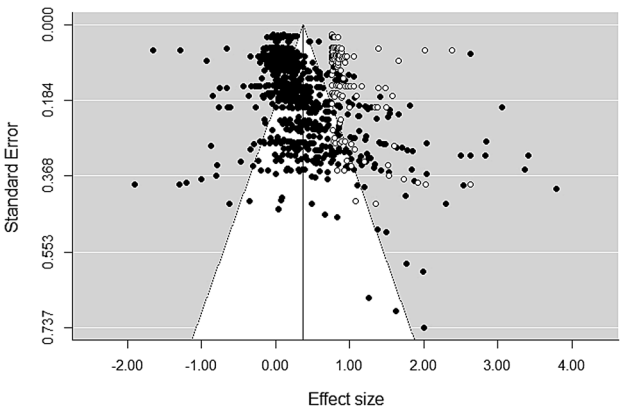

Fig. 2 Funnel plot

Secondary Analyses: Effects of Individual Training Components on Internalizing and Externalizing Behavior and Peer Relationship Problems

In a set of secondary analyses, we assessed whether individual training components moderated the effects on secondary

outcome measures. As only one included SST program did not include a cognitive-emotional component, this component category was not considered in the secondary analyses. SST program effects on peer relationship problems were not significantly associated with the inclusion of psychoeducation in the program, F (1, 54) = .248, p = .558, psychophysical components, F (1, 54) = .850, p = .361, or skill-building components, F (1, 54) = .009, p = .924. SST program effects on internalizing problem behavior were not significantly associated with the inclusion of psychoeducation, F (1, 180) = .158, p = .691, psychophysical components, F (1, 180) = .080, p = .778, or skill-building components, F (1, 180) = 2.162, p = .143. Finally, SST program effects on externalizing problem behavior were also not associated with the inclusion of psychoeducation, F (1, 230) = .548, p = .460, psychophysical components, F (1, 230) = .825, p = .365, or skill-building components, F (1, 230) = .097, p = .756. The results of these analyses are presented in Online Appendix F to H.

We assessed the mean estimated effect size per outcome domain (i.e., interpersonal skills, emotional skills, peer relationship problems, internalizing problem behavior and externalizing problem behavior) for the types of SST programs separately (Online Appendix I, Table I.1). The analyses showed that SEL programs had a significant, positive effect on interpersonal skills, d = .290, 95% CI [.187, .393], p < .001, emotional skills, d = .249, 95% CI [.142, .355], p < .001, and internalizing behavior problems, d = .128, 95% CI [.065, .191], p < .001. SEL programs did not have a significant effect on peer relationship problems, d = .171, 95% CI [.000, .342], p = .050, or externalizing problem behavior, d = .095, 95% CI [− .007, .197], p = .067. Programs that target (social) anxiety were effective in increasing emotional skills, d = .264, 95% CI [.082, .447], p < .05, and decreasing internalizing behavior problems, d = .384, 95% CI [.134, .634], p < .01, but not in increasing interpersonal skills, d = .259, 95% CI [− .248, .766], p = .203, or decreasing externalizing behavior problems, d = .402, 95% CI [-1.852, 2.692], p = .256. Programs that target resilience and self-esteem were only effective in increasing emotional skills, d = .287, 95% CI [.015, .559], p < .05, and decreasing internalizing problem behavior, d = .112, 95% CI [.046, .178], p < .01, but not in increasing interpersonal skills, d = − .006, 95% CI [− .094, .083], p = .893, or decreasing externalizing problem behavior, d = .165, 95% CI [− .795, 1.125], p = .273. Interestingly, anti-bullying programs showed the largest effects on emotional skills, d = .864, 95% CI [.742, .987], p < .001, but were not effective in reducing peer relationship problems, d = .666, 95% CI [− .010, .1.342], p = .053. Anti-bullying programs were also effective in increasing interpersonal skills, d = .709, 95% CI [.367, 1.050], p < .001, and decreasing internalizing and externalizing problem behavior, d = .846, 95% CI [.583, 1.110], p < .001 and d = .774, 95% CI [.099, .1.448], p < .05 , respectively. Programs that target prosocial behavior only had a significant, positive effect on interpersonal skills, d = .660, 95% CI [.273, 1.048], p < .01, and internalizing behavior problems, d = .198, 95% CI [.028, .369], p < .05. Programs aimed at reducing disruptive behavior were effective in increasing interpersonal skills, d = .253, 95% CI [.127, .378], p < .001, reducing internalizing behavior, d = .348, 95% CI [.092, .603], p < .05, and reducing externalizing behavior, d = .245, 95% CI [.086, .405], p < .01. These programs were not effective in improving emotional skills, d = .494, 95% CI [− .316, 1.304], p = .166, or reducing peer relationship problems, d = .219, 95% CI [− .740, 1.179], p = .429.

Discussion

Previous meta-analyses have reported on the general effectiveness of SST programs, but have not assessed if distinct training components can be related to SST program effects. The present meta-analysis related individual training components from as many as 60 SST programs to 839 effect sizes using a multilevel meta-analysis approach.

Our main focus was on the effects of SST programs on interpersonal and emotional skills in nonclinical children and adolescents, as a deficit in these outcomes can be considered a risk factor for problem behavior. Our findings demonstrate that SST programs have a significant, small positive effect on children's and adolescents' interpersonal and emotional skills. Next, we assessed whether the inclusion of specific training components was related to larger SST program effects on interpersonal and emotional skills. These analyses demonstrated that SST programs yield larger effects when psychoeducational exercises are included in the program. This finding indicates that it is necessary to educate children and adolescents on the skills SST programs aim to develop to increase the effect of SST programs on interpersonal and emotional skills. Notably, the inclusion of psychoeducational exercises had a curvilinear effect on SST program effects, whereby the strongest effects were found when three to six psychoeducational exercises were included in the program. A similar curvilinear effect was found for the number of skill-building exercises included in SST programs. The inclusion of 10 to 20 skill-building exercises was related to larger program effects on interpersonal and emotional skills, whereas the inclusion of more or fewer skill-building exercises seemed to hamper SST program effects. For the other two components (i.e., psychophysical components and cognitive-emotional components), we did not find an effect for the overall inclusion of the components nor a curvilinear effect for a specific number of exercises. None of the booster components were associated with larger SST program effects on interpersonal and emotional skills.

Programs that lasted more than 27 weeks yielded inferior results compared to programs of 10 to 16 weeks, which could indicate that the duration of an SST program is related to program effects. There might also be a curvilinear relationship between the duration of SST programs and their effects on children's and adolescents' interpersonal and emotional skills. After a certain point, more time spent on a program does not lead to superior results. The effectiveness of longer programs may be impaired by a difficulty to adhere to a protocol for an extended time (Lane et al. 2010). We could not assess whether this explanation holds true, however, because most studies did not report on the implementation quality of the SST programs, and thus, this was not coded. Controlling for implementation quality in future research might help to shed more light on this issue. Another possible explanation for the observed curvilinear relationship is that if a program does not match the level of participants' deficit in interpersonal and emotional skills or participants' treatment motivation, the program could be ineffective or have adverse effects (Bonta and Andrews 2007; Wilson and Hoge 2013).

Assessment of the overall effect size of SST programs on more distal outcome domains showed that SST programs also yield positive effects on peer relationship problems, internalizing behavior, and externalizing behavior. For these outcome domains, we did not find meaningful associations between the inclusion of psychoeducational, psychophysical, and skill-building components and SST program effects. Peer relationship problems, internalizing problem behavior, and externalizing problem behavior probably each have a wider range of contributing factors, which makes it more difficult to reveal the influence of independent training components.

The overall effect size found in this study is similar to the effects found in some previous meta-analyses (i.e., Durlak et al. 2011; Sklad et al. 2012) and somewhat smaller than the effects found in others (i.e., Denham and Almeida 1987; Schneider, 1992). Our study shows that the average person scores 13% higher on a social-emotional outcome after an SST program (Coe 2002). It is important to consider here that most of the included programs are universal prevention programs aimed at the general school population. One can assume that not all children participating in a universal program actually need the extra support, and in that context, a small positive effect for SST programs could mask the existence of intervention response subgroups (Vacha-Haase and Thompson 2004). For instance, a recent study on the effects of the Incredible Years parenting intervention uncovered that in a prevention context, only 18% of the targeted families benefitted from the program—but did so to a large extent (Cohen's $d = 1.45$)—whereas most families did not benefit. The families that benefitted the most were the ones with the highest parenting and child behavior problems (van Aar et al. 2019). Similarly, the small effect size on interpersonal and emotional skills found for SST programs in our study might indicate that a small group of children benefits substantially from the program, while a larger majority does not, simply because they already have adequate social skills. Heterogeneity in SST program effects across individuals points to the importance of analyzing subgroups of SST participants to gain insight into what training components work best for whom.

What do these findings tell us about the origin of SST program effects? Psychoeducational and skill-building exercises seem to be important for strong program effects, granted that the dosage is right. However, the inclusion of psychophysical and cognitive-emotional components does not necessarily seem to lead to superior SST program effects. This could signal that, to some extent, program effects are explained by the specific components in an SST program, which is in line with the specificity hypothesis (Chorpita and Daleiden 2009). Heterogeneity in SST program effects that is not explained by the inclusion of specific training components can perhaps, for a substantial part, be accounted for by several non-specific or "common factors." According to the common factors approach (Wampold et al. 1997), the alliance between a client and therapist, a therapist's belief in a program's effectiveness, and other therapist effects may be equally important for positive program effects (Messer and Wampold 2002). Such non-specific factors presumably also account for the variance between studies and programs. For example, meta-analyses (Horvath and Symonds 1991; Shirk and Karver 2003) have shown that a positive therapist–client relationship is related to better training outcomes. This might also be related to a therapists' experience (Mallinckrodt and Nelson 1991) and their expression of confidence and interest in clients (Saunders et al. 1989). Information about such common factors is generally not included in SST program manuals or the studies evaluating these programs. Thus, it was not possible to assess whether non-specific factors were related to SST program effects in this study. To be able to draw conclusions concerning the effectiveness of common factors in the future, it is important that information about non-specific factors such as a trainers' confidence in an SST program is reported in forthcoming studies.

The current meta-analysis provided some additional remarkable findings. For one, we found that SST programs provided by mental health professionals and SST programs given by school personnel (i.e., teachers or school nurses) yielded similar effects. The fact that we excluded studies evaluating SST program effects for special populations such as children with ADHD or autism may explain this finding. In the included samples, participants' problem behavior is mostly below the clinical cut-off, and the implementation of SST programs in such samples does not seem to require the expertise of mental health

professionals. Another explanation may be that consideration of the development of interpersonal and emotional skills has become more embedded in the tasks of school personnel, rendering them more qualified to provide SST programs and eliminating the necessity for specialized certification before program implementation. This could imply that a broader range of at-risk children and adolescents could be reached by including SST programs in the regular academic curriculum.

Checking the robustness of our main findings suggested that study quality impacts the magnitude of effects of SST programs. We found that studies with smaller sample sizes yielded larger effect sizes, which is likely related to the finding for study quality. Studies with a weak- or moderate-quality rating predominantly had a small sample size, whereas studies with a high-quality rating mostly had a large sample size. The absence of thorough randomization into experimental groups in low- and moderate-quality studies might threaten internal validity, leading to an overestimation of SST program effects, and this is mostly the case in studies with a small sample size (Weisburd et al. 2001). This finding is in line with findings from previous meta-analyses (e.g., Zhang et al. 2013).

A common disadvantage of meta-analyses is the dependence on the information available (Borenstein et al. 2009). In this study, we depended on study authors to provide us with program manuals. It is conceivable that authors were less likely to share a program that proved ineffective, which may have resulted in the significant publication bias we found, leading to an overestimation of program effects. Moreover, we could not investigate the interaction between components and have, therefore, assessed the training components independently. For this reason, we cannot make inferences about combinations of components that might perhaps amplify or attenuate SST program effects (Dusseldorp et al. 2013).

The agreement between coders might be considered a limitation of this study. It proved challenging to reach perfect alignment between coders concerning the content of the psychoeducational components of SST programs. We traced this to difficulties in the distinction between introductory information to SST exercises and pure psychoeducational exercises. This should be a point of attention if this study were to be replicated in the future.

Finally, we need to acknowledge the problem of multiplicity in this meta-analysis. We did not control for repeated significance testing, and consequently, some of our findings may be attributable to Type I (i.e., false-positive) error. Since this is a first meta-analysis investigating individual training components, more research is needed to draw strong conclusions on the effectiveness of individual SST components. The current findings can be used to generate hypotheses that might be tested in forthcoming studies, for example, using micro-trials (Howe et al. 2010) or additional meta-analyses.

Notwithstanding these limitations, our meta-analysis is the first to relate distinct training components to the effects of a large number of SST programs. We coded program information directly from program manuals, which allowed for detailed, schematic coding of the SST programs. Additionally, we assessed the dosage of training components, which provided valuable results we would have otherwise missed. The findings from the current study should be viewed as a first step towards uncovering the effective training components of SST programs.

In terms of practical implications, our findings suggest that SST programs yield positive effects on various outcome domains, and interpersonal and emotional skills seem to be most positively influenced. SST programs that include psychoeducational and skill-building exercises produce significantly stronger effects on interpersonal and emotional skills, but only when administered with the right dosage. When aiming to improve interpersonal and emotional skills, programs up to 16 weeks seem to generate optimal results. Booster components do not seem to influence program effects. These exercises could be replaced by more effective components or be removed to shorten programs. Such adjustments could make SST programs more cost-effective and leave more time for the regular academic curriculum, which is usually offered parallel to SST programs. However, it is also possible that the association between booster components and SST program effects only emerge after a follow-up period. Therefore, future meta-analyses of this type should also include long-term outcomes of SST program evaluations.

Nevertheless, the effect sizes found are moderate at best, and there is significant heterogeneity between studies, which suggests that not all participants benefit equally from SST programs. It is also important to keep in mind that the current meta-analysis only considered the immediate effects of SST programs. Currently, it is unclear whether SST program effects are generally sustained (e.g., Clarke et al. 2014) or lost at follow-up (e.g., Berry et al. 2016), or whether there are sleeper effects (e.g., Essau et al. 2012). An important next step in effective components research is assessing the synergistic effects of combinations of individual training components. Additionally, future research could meta-analyze individual participant data from pooled intervention datasets to further examine what works best for whom (Riley et al. 2010).

Funding Funding was provided by ZonMw (Grant No. "Effectief werken in de Jeugdsector," Project 729300011).

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical Approval This article does not contain any studies with human participants performed by any of the authors.

References

American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: American Psychological Association.

Ang, R. P. (2003). Social problem-solving skills training: Does it really work? *Child Care in Practice, 9*(1), 5–17. https://doi.org/10.1080/1357527031000113047.

Assink, M., & Wibbelink, C. J. M. (2016). Fitting three-level meta-analytic models in R: A step-by-step tutorial. *The Quantitative Methods for Psychology, 12*(3), 154–174. https://doi.org/10.20982/tqmp.12.3.p154.

Bandura, A. (1978). Social learning theory of aggression. *Journal of Communication, 28*(3), 12–29.

Berry, V., Axford, N., Blower, S., Taylor, R. S., Edwards, R. T., Tobin, K., et al. (2016). The effectiveness and micro-costing analysis of a universal, school-based, social–emotional learning programme in the UK: A cluster-randomised controlled trial. *School Mental Health, 8*(2), 238–256. https://doi.org/10.1007/s12310-015-9160-1.

Bonta, J., & Andrews, D. A. (2007). Risk-need-responsivity model for offender assessment and rehabilitation. *Rehabilitation*, *6*(1), 1–22.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. West Sussex: Wiley.

Caprara, G. V., Barbaranelli, C., Pastorelli, C., Bandura, A., & Zimbardo, P. G. (2000). Prosocial foundations of children's academic achievement. *Psychological Science, 11*(4), 302–306. https://doi.org/10.1111/1467-9280.00260.

Chorpita, B. F., & Daleiden, E. L. (2009). Mapping evidence-based treatments for children and adolescents: Application of the distillation and matching model to 615 treatments from 322 randomized trials. *Journal of Consulting and Clinical Psychology, 77*(3), 566–579. https://doi.org/10.1037/a0014565.

Clarke, A. M., Bunting, B., & Barry, M. M. (2014). Evaluating the implementation of a school-based emotional well-being programme: A cluster randomized controlled trial of Zippy's Friends for children in disadvantaged primary schools. *Health Education Research, 29*(5), 786–798. https://doi.org/10.1093/her/cyu047.

Coe, R. (September 2002). *It's the effect size, stupid. What effect size is and why it is important.* Paper presented at the Annual Conference of the British Educational Research Association, University of Exeter, England. Retrieved November 13, 2018, from https://www.leeds.ac.uk/educol/documents/00002182.htm.

Cook, C. R., Williams, K. R., Guerra, N. G., Kim, T. E., & Sadek, S. (2010). Predictors of bullying and victimization in childhood and adolescence: A meta-analytic investigation. *School Psychology Quarterly, 25*(2), 65–83. https://doi.org/10.1037/a0020149.

Denham, S. A., & Almeida, M. C. (1987). Children's social problem-solving skills, behavioral adjustment, and interventions: A meta-analysis evaluating theory and practice. *Journal of Applied Developmental Psychology, 8*(4), 391–409. https://doi.org/10.1016/0193-3973(87)90029-3.

Diekstra, R. F. W. (2008). Effectiveness of school-based social and emotional education programmes worldwide. In C. Clouder, B. Dahlin, R. Diekstra, P. F. Berrocal, B. Heys, L. Lantieri, & H. Paschen (Eds.), *Social and emotional education: An International Analysis* (pp. 255–312). Fundacion Marcelino Botin: Santander.

Dishion, T. J., McCord, J., & Poulin, F. (1999). When interventions harm: Peer groups and problem behavior. *American Psychologist, 54*(9), 755. https://doi.org/10.1037/0003-066X.54.9.755.

Durlak, J. A. (2009). How to select, calculate, and interpret effect sizes. *Journal of Pediatric Psychology*, *34*(9), 917–928.

Durlak, J. A., & Lipsey, M. W. (1991). A practitioner's guide to meta-analysis. *American Journal of Community Psychology, 19*(3), 291–332. https://doi.org/10.1007/BF00938026.

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development, 82*(1), 405–432. https://doi.org/10.1111/j.1467-8624.2010.01564.x.

Dusseldorp, E., van Genugten, L., van Buuren, S., Verheijden, M. W., & van Empelen, P. (2013). Combinations of techniques that effectively change health behavior: Evidence from Meta-CART analysis. *Health Psychology, 33*(12), 1530–1540. https://doi.org/10.1037/hea0000018.

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56,* 455–463.

Effective Public Health Practice Project. (1998). Dictionary for the Effective Public Health Practice Project. Retrieved August 3, 2016, from https://merst.ca/wp-content/uploads/2018/02/quality-assessment-tool_2010.pdf.

Essau, C. A., Conradt, J., Sasagawa, S., & Ollendick, T. H. (2012). Prevention of anxiety symptoms in children: Results from a universal school-based trial. *Behavior Therapy, 43*(2), 450–464. https://doi.org/10.1016/j.beth.2011.08.003.

Frey, K. S., Hirschstein, M. K., & Guzzo, B. A. (2015). Second step: Preventing aggression by promoting social competence. *Journal of Emotional and Behavioral Disorders, 8*(2), 102–112.

Gresham, F. M., & Elliot, S. N. (1987). The relationship between adaptive behavior and social skills: Issues in definition and assessment. *The Journal of Special Education, 21*(1), 167–181. https://doi.org/10.1177/002246698702100115.

Horvath, A. O., & Symonds, B. D. (1991). Relation between working alliance and outcome in psychotherapy: A meta-analysis. *Journal of Counseling Psychology, 38*(2), 139–149. https://doi.org/10.1037/0022-0167.38.2.139.

Howe, G. W., Beach, S. R. H., & Brody, G. H. (2010). Microtrial methods for translating gene-environment dynamics into preventive interventions. *Prevention Science, 11,* 343–354. https://doi.org/10.1007/s11121-010-0177-2.

Kavale, K. A., & Forness, S. R. (1996). Social skills deficits and learning disabilities: A meta -analysis. *Journal of Learning Disabilities, 29*(3), 226–237. https://doi.org/10.1177/002221949602900301.

Knapp, G., & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine, 22,* 2693–2710. https://doi.org/10.1002/sim.1482.

Ladd, G. W., & Mize, J. (1983). A cognitive-social learning model of social-skill training. *Psychological Review, 90*(2), 127–157.

Lane, K. L., Bocian, K. M., MacMillan, D. L., & Gresham, F. M. (2010). Treatment integrity: An essential–but often forgotten–component of school-based interventions. *Preventing School Failure: Alternative Education for Children and Youth, 48*(3), 36–43.

Leijten, P., Dishion, T. J., Thomaes, S., Raaijmakers, M. A. J., Orobio de Castro, B., & Matthys, W. (2015). Bringing parenting interventions back to the future: How randomized microtrials may benefit parenting intervention efficacy. *Clinical Psychology: Science and Practice, 22*(1), 47–57. https://doi.org/10.1111/cpsp.12087.

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., et al. (2009). The PRISMA statement for

reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *Journal of Clinical Epidemiology, 62,* 1–34. https://doi.org/10.1016/j.jclinepi.2009.06.006.

Lösel, F., & Beelmann, A. (2003). Effects of child skills training in preventing antisocial behavior: A systematic review of randomized evaluations. *The ANNALS of the American Academy of Political and Social Science, 587*(1), 84–109. https://doi.org/10.1177/0002716202250793.

Malecki, C. K., & Elliot, S. N. (2002). Children's social behaviors as predictors of academic achievement: A longitudinal analysis. *School Psychology Quarterly, 17*(1), 1–23. https://doi.org/10.1521/scpq.17.1.1.19902.

Mallinckrodt, B., & Nelson, M. L. (1991). Counselor training level and the formation of the psychotherapeutic working alliance. *Journal of Counseling Psychology, 38*(2), 133–138. https://doi.org/10.1037/0022-0167.38.2.133.

McCartney, K., & Rosenthal, R. (2000). Effect size, practical importance, and social policy for children. *Child Development, 71*(1), 173–180.

McCelland, M. M., & Scalzo, C. (2006). Social skills deficits. In M. Hersen (Ed.), *Clinician's handbook of child behavioural assessment* (pp. 313–335). Amsterdam: Elsevier.

Messer, S. B., & Wampold, B. E. (2002). Let's face facts: Common factors are more potent than specific therapy ingredients. *Clinical Psychology: Science and Practice, 9*(1), 21–25. https://doi.org/10.1093/clipsy/9.1.21.

Michie, S., Richardson, M., Johnston, M., Abraham, C., Francis, J., Hardeman, W., et al. (2013). The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: Building an international consensus for the reporting of behavior change interventions. *Annals of Behavioral Medicine, 46*(1), 89–95. https://doi.org/10.1007/s12160-013-9486-6.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & the PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine, 151*(4), 264–269.

Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods, 11*(2), 364–386. https://doi.org/10.1177/1094428106291059.

Reddy, L. A., Newman, E., De Thomas, C. A., & Chun, V. (2009). Effectiveness of school-based prevention and intervention programs for children and adolescents with emotional disturbance: A meta-analysis. *Journal of School Psychology, 47*(2), 77–99. https://doi.org/10.1016/j.jsp.2008.11.001.

Riley, R. D., Lambert, P. C., & Abo-Zaid, G. (2010). Meta-analysis of individual participant data: Rationale, conduct, and reporting. *British Medical Journal, 340,* 521–525. https://doi.org/10.1136/bmj.c221.

Saunders, S. M., Howard, K. I., & Orlinsky, D. E. (1989). The Therapeutic Bond Scales: Psychometric characteristics and relationship to treatment effectiveness. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1*(4), 323–330. https://doi.org/10.1037/1040-3590.1.4.323.

Schneider, B. H. (1992). Didactic methods for enhancing children's peer relations: A quantitative review. *Clinical Psychology Review, 12*(3), 363–382. https://doi.org/10.1016/0272-7358(92)90142-U.

Segrin, C. (2000). Social skills deficits associated with depression. *Clinical Psychology Review, 20*(3), 379–403. https://doi.org/10.1016/S0272-7358(98)00104-4.

Shirk, S. R., & Karver, M. (2003). Prediction of treatment outcome from relationship variables in child and adolescent therapy: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 71*(3), 452–464. https://doi.org/10.1037/0022-006X.71.3.452.

Skinner, B. F. (1953). *Science and human behavior*. New York: Simon and Schuster. https://doi.org/10.1037/h0052427.

Sklad, M., Diekstra, R., de Ritter, M., Ben, J., & Gravestijn, C. (2012). Effectiveness of school-based universal social, emotional, and behavioural programs: Do they enhance students' development in the area of skill, behaviour, and adjustment? *Psychology in the Schools, 49*(9), 153–178. https://doi.org/10.1002/pits.21641.

Spence, S. H. (2003). Social skills training with children and young people: Theory, evidence and practice. *Child and Adolescent Mental Health, 8*(2), 84–96. https://doi.org/10.1111/1475-3588.00051.

Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods, 5*(1), 60–78.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics*. New York: Pearson.

Taylor, R. D., Oberle, E., Durlak, J. A., & Weissberg, R. P. (2017). Promoting positive youth development through school-based social and emotional learning interventions: A meta-analysis of follow-up effects. *Child Development, 88*(4), 1156–1171. https://doi.org/10.1111/cdev.12864.

Thomas, B. H., Ciliska, D., Dobbins, M., & Micucci, S. (2004). A process for systematically reviewing the literature: Providing the research evidence for public health nursing interventions. *Worldviews on Evidence-Based Nursing, 1*(3), 176–184.

Vacha-Haase, T., & Thompson, B. (2004). How to estimate and interpret various effect sizes. *Journal of Counseling Psychology, 51*(4), 473–481. https://doi.org/10.1037/0022-0167.51.4.473.

van Aar, J., Leijten, P., Orobio de Castro, B., Weeland, J., Matthys, W., Chhangur, R., et al. (2019). Families who benefit and families who do not: Integrating person- and variable-centered analyses of parenting intervention responses. *Journal of the American Academy of Child and Adolescent Psychiatry, 58,* 993–1003e1.

van den Noortgate, W., & Onghena, P. (2003). Multilevel meta-analysis: A comparison with traditional meta-analytical procedures. *Educational and Psychological Measurement, 63*(5), 765–790. https://doi.org/10.1177/0013164402251027.

Veerman, J. W., Spanjaard, H., & Yperen, T. (2015). Elementen die er toe doen! Werkdocument voor een taxonomie van werkzame elementen in de zorg voor jeugd. ZonMw, Effectief werken in de Jeugdzorg. Retrieved July 3, 2017, from https://publicaties.zonmw.nl/werkdocument-voor-een-taxonomie-van-werkzame-elementen-in-de-zorg-voor-jeugd/.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistic Software, 36*(3), 1–48.

Wampold, B. E., Mondin, G. W., Moody, M., Stich, F., Benson, K., & Ahn, H. (1997). A meta-analysis of outcome studies comparing bona fide psychotherapies: Empirically, "all must have prizes". *Psychological Bulletin, 122*(3), 203–215. https://doi.org/10.1037/0033-2909.122.3.203.

Weisburd, D., Lum, C. M., & Petrosino, A. (2001). Does research design affect study outcomes in criminal justice? *The Annals of the American Academy of Political and Social Science, 578,* 50–70. https://doi.org/10.1177/000271620157800104.

Wilson, H. A., & Hoge, R. B. (2013). The effect of youth diversion programs on recidivism: A meta-analytic review. *Criminal Justice and Behavior, 40*(5), 497–518. https://doi.org/10.1177/0093854812451089.

Zhang, Z., Xu, X., & Ni, H. (2013). Small studies may overestimate the effect sizes in critical care meta-analyses: A meta-epidemiological study. *Critical Care, 17*(1), 1–9.